

Into my Packet



後藤滋樹の

## 新・社会楽

後藤滋樹  
goto@goto.info.waseda.ac.jp  
早稲田大学 理工学部 情報学科

## 第53回「不思議な法則」

## 【甦るジップの法則】

今は昔、ジップという人が長い小説（ジェームズ・ジョイスの「ユリシーズ」）の260,430個の英単語といくつかの新聞記事の43,989個の英単語の出現頻度を調べた。英単語を出現の多い順に並べると、(1) the、(2) of、(3) and、(4) toとなる。さらに興味深いことには、第1位のtheは全体の出現数の約10%を占める。第2位のofは5%、第3位のandは3.3%という具合になる。つまり式で書くと次のようになるという。

$$(\text{出現率}\%) = \frac{10}{(\text{順位})}$$

このジップの法則は、英単語だけでなく、あらゆるところに適用できる。

## 【都市の人口、インターネットのホスト数】

ピアース先生の著書<sup>①</sup>によると、ジップの法則は多くの国の都市の人口とその大きさの順位にもあてはまるという。

インターネット時代の人口に相当するのは、接続されているホスト数である。さっそくNetwork Wizards Inc. (<http://www.nw.com/>) の統計を用いて、ドメイン別のホスト数を調べてみよう。1999年の1月現在のインターネットの統計によると、ドメイン別のホスト数は上位から順番に(1) comが12,140,747台、(2) netが8,856,687台、(3) eduが5,022,815台、(4) jpが1,687,534台である。その後は(5) usが1,562,391台、(6) milが1,510,440台、(7) ukが1,423,804台、(8) deが1,316,893台、(9) caが1,119,172台、(10) auが792,351台と続く。

ホスト数の場合には、英単語のジップの法則と全く同じにはならないが、その傾向は大変によく似ている。ジップの法則を対数目盛のグラフに書くと右下がりの直線になるのだが、インターネットのホスト数のグラフも実際に同じ傾向を示す。

## 【人気のあるウェブへのアクセス数の頻度】

インターネットの世界でジップの法則を適用する試みは、実は以前から提案されている。その代表的な研究は、ウェブのアクセス頻度である。人気のあるページを第1位から順番に並べる。そ

のヒット数の統計はジップの法則に従うという。

このような研究は国内にも多数ある。ここでは海外の資料として<http://www.useit.com/alertbox/zipf.html>を参照する。この研究の中では<http://www.sun.com/>（サン・マイクロシステムズのウェブサイト）の各ページの1か月間のアクセスの統計を分析しており、人気の順位とヒット数がほぼジップの法則に従うとしている。

私の研究室でも、学生と一緒に早稲田大学内のキャッシュサーバーのログを分析してみた。上のような単一のサーバー内だけでなく、多数のサーバーを相手とするウェブのリクエストでも、人気順位と頻度数がジップの法則に合致するという結論を得ている。



## 【Emacsの編集コマンド】

コンピュータを使用する場面においてもジップの法則が活躍する。筆者の友人の奥乃博氏（東京理科大教授）は、UNIX環境のテキストエディターであるEmacsを利用する人のコマンドの履歴を大量に分析した。

結論は、Emacsのコマンドの使用頻度もジップの法則に従うという<sup>②</sup>。そういえば、私はMuleというテキストエディターを用いて原稿を書いている。MuleもEmacsの一種であるから、私の執筆もジップの法則に従っているのだ。

ほかの分野の例として映画の観客動員数を調べてみた。<http://www.fm-kyoto.co.jp/>に京都エリアの映画館の情報がある。98年12月23日のデータでは、第1位の「アルマゲドン」の観客動員数はおよそ4,300人、第2位の「ジョー・ブラックをよろしく」はおよそ2,700人、そして第3位の「6デイズ/7ナイツ」はおよそ1,500人とある。第2位が第1位の半分の人数、第3位が1/3の人数となるのはジップの法則にほかならない。

このように書いてくると、ジップの法則は万能のように思うかもしれない。もちろん、そんなことはないのだが、インターネットの時代になって各種の統計が取り易くなり、昔ながらの法則が再び脚光を浴びているのは面白いことである。

① J.R.ピアース著、鎮目義夫訳「記号・シグナル・ノイズ」白揚社1988。  
上の原著はJohn R. Pierce, An Introduction to Information Theory: Symbols, Signals and Noise, Second, Revised Edition, Dover, 1980.

② 奥乃博「画面エディタEmacsのユーザ特性について」  
情報処理学会 第25回プログラミングシンポジウム1984.



## [インターネットマガジン バックナンバーアーカイブ] ご利用上の注意

このPDFファイルは、株式会社インプレスR&D(株式会社インプレスから分割)が1994年～2006年まで発行した月刊誌『インターネットマガジン』の誌面をPDF化し、「インターネットマガジン バックナンバーアーカイブ」として以下のウェブサイト「All-in-One INTERNET magazine 2.0」で公開しているものです。

<http://i.impressRD.jp/bn>

このファイルをご利用いただくにあたり、下記の注意事項を必ずお読みください。

- 記載されている内容(技術解説、URL、団体・企業名、商品名、価格、プレゼント募集、アンケートなど)は発行当時のものです。
- 収録されている内容は著作権法上の保護を受けています。著作権はそれぞれの記事の著作者(執筆者、写真の撮影者、イラストの作成者、編集部など)が保持しています。
- 著作者から許諾が得られなかった著作物は収録されていない場合があります。
- このファイルやその内容を改変したり、商用を目的として再利用することはできません。あくまで個人や企業の非商用利用での閲覧、複製、送信に限られます。
- 収録されている内容を何らかの媒体に引用としてご利用する際は、出典として媒体名および月号、該当ページ番号、発行元(株式会社インプレス R&D)、コピーライトなどの情報をご明記ください。
- オリジナルの雑誌の発行時点では、株式会社インプレス R&D(当時は株式会社インプレス)と著作権者は内容が正確なものであるように最大限に努めましたが、すべての情報が完全に正確であることは保証できません。このファイルの内容に起因する直接のおよび間接的な損害に対して、一切の責任を負いません。お客様個人の責任においてご利用ください。

このファイルに関するお問い合わせ先

**株式会社インプレスR&D**

All-in-One INTERNET magazine 編集部

[im-info@impress.co.jp](mailto:im-info@impress.co.jp)